

```

<document index="2" page="1">
  <title>Andreescu, Emilia - Cum Si Cat Trebuie Sa Manance ...</title>
  <url>https://www.scribd.com/document/1073006320/Andreescu-Emilia-Cum-Si-Cat-Trebuie-Sa-
Manance-Copilul</url>
  <id>1073006320</id>
</document>
<document index="3" page="1">
  <title>Russu, George - Boli Si Sindroame Cu Nume Proprii ...</title>
  <url>https://www.scribd.com/document/1073006317/Russu-George-Boli-Si-Sindroame-Cu-Nume-
Proprii-Volumul-2</url>
  <id>1073006317</id>
</document>
<document index="4" page="1">
  <title>Badea, Mariana - Bumerang de Manele Lingvistice</title>
  <url>https://www.scribd.com/document/1073006316/Badea-Mariana-Bumerang-de-Manele-
Lingvistice</url>
  <id>1073006316</id>
</document>
<document index="5" page="1">
  <title>Burca, Alina - Sincronie Si Diacronie in Sintaxa S...</title>
  <url>https://www.scribd.com/document/1073006315/Burca-Alina-Sincronie-Si-Diacronie-in-
Sintaxa-Si-Stilistica-Adjectivului</url>
  <id>1073006315</id>
</document>

```

Sarcina este ca formula regex pentru notepad++ sa faca un replace ca apoi sa ramana doar asa:

```

https://www.scribd.com/document/1073006320/Andreescu-Emilia-Cum-Si-Cat-Trebuie-Sa-Manance-
Copilul
https://www.scribd.com/document/1073006317/Russu-George-Boli-Si-Sindroame-Cu-Nume-Proprii-
Volumul-2
https://www.scribd.com/document/1073006316/Badea-Mariana-Bumerang-de-Manele-Lingvistice
https://www.scribd.com/document/1073006315/Burca-Alina-Sincronie-Si-Diacronie-in-Sintaxa-Si-
Stilistica-Adjectivului

```

REZOLVARE:

Iată soluția simplă pentru Notepad++ folosind funcția Find & Replace (Caută și Înlocuiește).

Pattern-ul Regex

1. Deschide documentul în Notepad++ și apasă Ctrl + H (Înlocuire).
2. La Find what (De căutat), introdu:

```
(?s)<document.*?<url>(.*?)</url>.*?</document>\s*
```

3. La Replace with (Înlocuiește cu), introdu:

```
$1\r\n
```

4. Jos, la Search Mode (Mod de căutare), bifează Regular expression (Expresie regulată).
5. Apasă pe Replace All (Înlocuiește tot).

Cum funcționează:

- * (?s) – Îi spune programului că textul se întinde pe mai multe rânduri (dot matches newline).
- * <url>(.*?)</url> – Găsește link-ul și îl salvează în prima „grupă” de memorie.
- * \$1\r\n – Șterge tot codul din jur și păstrează doar link-ul salvat, adăugând un rând nou după fiecare.

Dacă vrei să modifizi și alte liste similare sau ai nevoie de ajutor cu curățarea altor etichete XML, lasă-mi un mesaj!

perfect. acum cu generic formulas din lista de mai jos:

GENERIC FORMULAS

Lista Generic Regex:

```
(?s)(?-i:BSR).*(?<=FR1)(.)*(?=FR2)(?=.)*?ESR)

(?:\G(?:!\A)|^.*?(?:BSR))?(?:(!ESR).)*?\K(?:FR) / REPLACE RR

(?:.*?BSR|\G).*?FR\K(.*)?(?=ESR.*?ESR)

(?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\KFR

(BSR)s*FR(.*)FR\s*(ESR) Replace by: \1\2\3

(?s-i)(?:BSR|(?!\A)\G)(?:(!ESR).)*?\K(FR)

(BSR.*?)(FR\s*ESR)

(?:BSR|\G).*\K(?:FR(*SKIP)(*FAIL))|ESR)

(?:BSR|\G).*\K(?:FR(*SKIP)(*FAIL))|<(?:(?!ESR).)+?>)

(?:BSR|\G).*\K(?:FR(*SKIP)(*FAIL))|<(?:(?!ESR1|ESR2).)+?>)

BSR(?:[^\<]+|<(?!FR1|FR2))?</p>(*SKIP)(FAIL)|<(?!FR1|FR2)[^\>]+>

BSR(?:[^\<]+|<(?!FR))?</p>(*SKIP)(FAIL)|<(?!FR)[^\>]+>

(?-s)(?-i:BSR|(?!\A)\G)(?s-i:.)*?ESR)(*SKIP)(*FAIL)|FR

(?s)(?-i:BSR|(?!\A)\G).*\K(!ESR)
```

(?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\x20\\K(FR)
 (?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\x20\\KFR(=?\x20)
 (?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\x20\\KFR
 (?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\x20\\KFR(=?\x20)
 (?s-i)(?:BSR|(?!\A)\G)(?:!(?!ESR).)*?K<(?!(?:FR1|FR2))[^>]*>
 (?s)(?-i:BSR|(?!\A)\G).*?K(?!FR1)(.*?)(?!FR2)(?!ESR)
 (?s)(?-i:BSR).*?(?!FR1)(.*?)(?!FR2)(?!ESR)
 BSR[\s\S]+FR1\s+(\S+)\s+FR2[\s\S]+ESR
 (?s)(?!BSR[\s\S]+FR1\s+(\S+)\s+FR2[\s\S]+ESR)
 (?s)BSR[\s\S]+FR1\s+(\S+)\s+FR2[\s\S]+ESR
 (?-s)(?-i:BSR|(?!\A)\G)(?s-i:.*?ESR)(*SKIP)(*FAIL)|FR
 (?-s)(?-i:BSR|(?!\A)\G).*?KFR(?!ESR)
 (?-s)(?-i:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?K(?-si:FR)
 (?-s)(?-i:BSR|(?!\A)\G).*?K(?-si:FR)(?!s-i:.*?ESR))
 (?s)BSR(FR)(?-s).+ESR or BSR(FR).(?!s:(?!ESR)))
 (?s)(?-i:BSR|(?!\A)\G).*?KFR(?!ESR)
 (?s)(?-i:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?K(?-si:FR)
 (?s)(?-i:BSR|(?!\A)\G).*?K(?-si:FR)(?!s-i:.*?ESR))
 (?s)BSR(FR)(?-s).+ESR or BSR(FR).(?!s:(?!ESR)))
 (?-si:BSR|(?!\A)\G).*?KFR(?!ESR)
 (?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?K(?-si:FR)
 (?-si:BSR|(?!\A)\G).*?K(?-si:FR)(?!s-i:.*?ESR))
 (?s)BSR(FR)(?-s).+ESR sau BSR(FR).(?!s:(?!ESR)))
 (?-s)(?-i:BSR|(?!\A)\G).*?K(?-i:FR)(?!ESR)

(?-si:BSR|(?!\A)\G)(?s-i:(?!ESR).)*?\K(FR)+ REPLACE BY: \2 sau \3
 (?-i:BSR|G(?!\^))(\G)(?s:(?!ESR).)*?\K(?-i:FR)
 BSR(?!\A)(?-s:.*?)\K(?-si:FR)(?=(?s-i:.*?ESR))
 BSR\K.*?(?!A)(?-s:.*?)\K(?-si:FR)(?=(?s-i:.*?ESR))
 (?-i:BSR|(?!\A)\G)(?s:(?!ESR).)*?\K(?-i:FR)
 (?-i:BSR|(?!\^)\G)(?s:(?!ESR).)*?\K(?-i:FR)
 (?s)\A.*BSR(FR)*ESR.*\z|\A.*BSR(FR)+ESR.*\z
 (?s)\A.*BSR(FR)+ESR.*\z
 (?s)\A.*BSR+(FR)+ESR.*\z
 (?:BSR|G)\S*\K(?:FR)(?=. *ESR)
 (?:BSR|G)\S*\K(FR)(?=. *ESR)
 (?:BSR|G)\S*\K(?:\h+FR)(?=. *ESR)
 (?<!=\x20)(?<!=)FR(?:>|ESR|\x20>|\x20/>|\?>|\x20\?>|BSR)
 (?<!=\x20)(?<!=)FR(?:>|ESR|\x20>|\x20/>|\?>|\x20\?>|\x20BSR)
 (?<!=\x20)(?<!=)FR(?:>|ESR|\x20\?>|\x20BSR)
 (?-is)^(?!.*BSR)(?!.*ESR)(?:(FR))
 (?-is)^(?!.*BSR)(?!.*ESR)(?:(FR)(.+))
 (BSR).*\K(?:\h+(FR))(?=. *ESR)
 (?-i:BSR|G(?!\^))(\G)(?s:(ESR).)*?\K(?-i:FR)
 (BSR)+(.)+\K(FR)(?s:(?=. *(ESR)))
 (BSR)(FR)(.*?)(ESR)
 (?s-i)^\.+BSR\R|ESR.+|(?-s)^(FR)((.+))\$
 (?s-i)^\.+BSR\R|ESR.+|(?-s)^(FR)((.+))\$
 (BSR)(.*?)(FR)(.*?)(ESR)
 (?s)(?-i:(BSR).+?ESR|G(?!\^))((?!ESR)).)*?\K(FR)

(?s)(?-i:BSR.+?ESR\|G(?!\^))((?!ESR).)*?\KFR

(BSR)+(.)+\K(FR)(?=. *ESR)

(BSR)+(.)+\K(FR)(?=.*)(ESR)

(BSR)*\K(?: (FR))(?=. *ESR)

(BSR)\K(?-s)(.*?)\KFR(?=. *ESR)

. *?(BSR)(?-s)(.*?)(FR).*(. *ESR)

(BSR).*(\b(FR)\b.*).*(ESR)

(?:BSR\|G). *?\K(?:FR(*SKIP)(*FAIL)|<(?!ESR)[^>]*>)

BSR(?=(?: (?!ESR).)*?FR).+?ESR\R?

BSR(?=(?: (?!FR).)*?FR).+?ESR\R?

BSR(?=(?: (?!ESR).)*?FR)(?=(?: (?!ESR).)*?FR).+?ESR\R?

BSR(?=(?: (?!ESR).)*?FR)(?=(?: (?!ESR).)).+?ESR\R?

(?s)(?:BSR\|G)(?: (?!ESR).)*?\K(FR)

(?s)(?:BSR\|G(?!\^))(?: (?!ESR).)*?\KFR

(?!BSR)*\K(?!BSR)(?: (FR)))(?=. *ESR)

(?s)(?:BSR\|G)(?: (?!ESR).)*?\K(?:FR)

(?s)(?:BSR|(?!\A)\G)(?: (?!ESR).)*?\KFR

(?s)(?:BSR\|G). *?\K(?:FR(*SKIP)(*FAIL)|<(?: (?!ESR).)+?>)

(?:BSR\|G). *?\K(?:FR(*SKIP)(*FAIL)|<(?: (?!ESR).)+?>)

(?:BSR\|G). *?\K). *?(FR)(?=. *ESR)

(?:BSR\|G). *?\K)(FR). *?(?=. *ESR)

(?:BSR\|G)(FR). *?\K). *?(?=. *ESR)

(?s-i)(?:BSR\|G)(?: (?!ESR).)*?\K(FR)

(?:BSR\|G)(FR). *?(FR)\K). *?(?=. *ESR)

(?s-i)(?:<p[^\>]*>|(?!\A)\G(?:!(?:</p>).)*?\K<(?!(?:/[ap]>|a\x20[^\>]+>))[\^\>]*>

(?s-i)(?:BSR|(?!\A)\G(?:!(?:ESR).)*?\K<(?!(?:FR1|FR2))[\^\>]*>

(?:BSR|\G).*\K(?:FR(*SKIP)(*FAIL)|<(?!ESR).)+?>)

(?:BSR|\G).*\K(?:FR(*SKIP)(*FAIL)|<(?!ESR1|ESR2).)+?>)

BSR(?:[^\<]+|<(?!FR1|FR2))?ESR(*SKIP)(FAIL)|<(?!FR1|FR2)[^\>]+>

SPECIFICATII:

Let BSR (Begin Search-region Regex) be the regex which defines the beginning of the area where the search for FR, must start

Let ESR (End Search-region Regex) be the regex which defines, implicitly, the area where the search for FR, must end

Let FR (Find Regex) be the regex which defines the char, string or expression to be searched

Let RR (Replacement Regex) be the regex which defines the char, string or expression which must replace the FR expression. REPLACE RR (means can be \1 \2 ...or \$1 \$2 or other

1. Componente cheie:

- BSR (Begin Search-region Regex): Marchează începutul zonei de căutare
- ESR (End Search-region Regex): Marchează sfârșitul zonei de căutare
- FR (Find Regex): Expresia de căutat în zona definită
- RR (Replacement Regex): Expresia cu care se înlocuiește FR (dacă e cazul)

Vreau cel puțin 2 expresii regulate și să îmi spui și care generic formula ai folosit, dar să îmi dai și formula înlocuită

Iată două formule generice selectate exact din lista ta, adaptate complet pentru a extrage doar URL-urile, împreună cu instrucțiunile precise de înlocuire (Replace) pentru fiecare.

Opțiunea 1: Folosind formula generică bazată pe 5 grupuri

Formula generică din listă:

(BSR)(.*?)(FR)(.*?)(ESR)

Adaptarea componentelor:

* BSR: <document[^\>]*>

* ESR: </document>\s*

* FR: <url>(https?:/[^\s<]+)</url> (am adăugat o paranteză rotundă în interiorul FR pentru a izola link-ul curat)

Setările pentru Notepad++:

* Find what: (<document[^\>]*>)(.*?)(<url>(https?:/[^\s<]+)</url>)(.*?)(</document>\s*)

* Replace with: \$4\r\n

* Search Mode: Regular expression (bifează și . matches newline)

De ce funcționează: Formula împarte textul în grupuri. Link-ul curat devine al 4-lea grup (\$4), iar înlocuirea șterge restul etichetelor XML și păstrează doar URL-ul urmat de un rând nou.

Iată o altă formulă din lista ta care lucrează corect și nativ cu blocuri multiple, fiindcă folosește o verificare strictă pentru a nu depăși sfârșitul blocului curent (ESR).

Opțiunea 2 Corectă: Folosind formula generică cu limitare de bloc

Formula generică din listă:

BSR(?=(?:(!ESR).)*?FR).+?ESR\R?

Această formulă este excelentă pentru blocuri multiple deoarece folosește .+? (potrivire leneșă/lazy) și se asigură prin tehnica lookahead (?!(ESR)) că nu sare peste granița etichetei de final.

Adaptarea componentelor pentru cazul tău:

* BSR: <document[^>]*>

* ESR: </document>

* FR (extins pentru captură): <url>(.*?)</url> (înlocuim partea interioară a blocului pentru a prinde direct grupul de care avem nevoie)

Setările pentru Notepad++:

* Find what (Caută):

(?s)<document[^>]*>(?(?:(!</document>).)*?<url>)(.*?)(?<url>https?:/[^\s<]+</url>).*?</document>\s*

* Replace with (Înlocuiește cu):

\$2\r\n

(Aici \$2 va extrage exact link-ul curat din interiorul grupului <url>, eliminând în siguranță restul textului din bloc fără să afecteze celelalte documente).

* Search Mode: Regular expression

Spune-mi dacă această variantă rulează acum corect pe structura ta sau dacă vrei să adaptăm una dintre formulele care folosesc \K (clauza de resetare) pentru a modifica doar o porțiune din interiorul link-urilor!